

aThis Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/57, 9/64, 1/21, 1/19, 5/10, A61K 38/48, C07K 16/40, C12Q 1/68	A2	(11) International Publication Number: WO 99/41387 (43) International Publication Date: 19 August 1999 (19.08.99)
(21) International Application Number: PCT/US99/02571 (22) International Filing Date: 5 February 1999 (05.02.99) (30) Priority Data: 09/025,059 17 February 1998 (17.02.98) US (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 09/025,059 (CIP) Filed on 17 February 1998 (17.02.98) (71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). (74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: PROSTATE-ASSOCIATED SERINE PROTEASE (57) Abstract The invention provides a human prostate-associated serine protease (PRASP) and polynucleotides which identify and encode PRASP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for treating or preventing disorders associated with expression of PRASP.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

PROSTATE-ASSOCIATED SERINE PROTEASE

TECHNICAL FIELD

5 This invention relates to nucleic acid and amino acid sequences of a prostate-associated serine protease and to the use of these sequences in the diagnosis, treatment, and prevention of reproductive disorders and cancer.

BACKGROUND OF THE INVENTION

10 Proteolysis is one of the most important and frequent enzymatic reactions that occur both within and outside of cells. Proteolysis results in the activation, degradation, and maturation of nascent polypeptides. These modifications ultimately influence cell cycle progression, cell-cell signaling, metabolism, and many other processes required for development and survival of multicellular organisms. Proteolytic enzymes, or proteases,
15 catalyze the hydrolytic cleavage of peptide bonds. Proteases can be classified based on their enzymatic mechanism, substrate specificity, active site configuration, and overall three-dimensional structure. The mammalian serine proteases comprise one of the most well-characterized and versatile protease families. This family is exemplified by the pancreatic and plasma serine proteases.

20 Pancreatic serine proteases are secreted from the pancreas into the duodenum where they degrade proteins ingested in food. Examples of these proteases include chymotrypsin, trypsin, elastase, and pancreatic kallikrein. Plasma serine proteases, which include thrombin and C1r, are involved in blood coagulation and immune response. Thrombin converts fibrinogen, a large soluble plasma protein, into fibrin, a smaller
25 insoluble protein that aggregates to form blood clots. C1r is a component of the complement system, a complex of proteins that perforates the cell membranes of invading microorganisms.

 Most mammalian serine proteases are synthesized as zymogens, inactive precursors that are activated by protease cascades. For example, trypsinogen is converted
30 to its active form, trypsin, by enteropeptidase. Enteropeptidase is an intestinal protease that removes an N-terminal fragment from trypsinogen. The remaining active fragment is

trypsin, which in turn activates the precursors of the other pancreatic enzymes. Likewise, proteolysis of prothrombin, the precursor of thrombin, generates three separate polypeptide fragments. The N-terminal fragment is released while the other two fragments, which comprise active thrombin, remain associated through disulfide bonds.

5 The catalytic active site is conserved among mammalian serine proteases. A defining characteristic of the active site is the catalytic triad, the conserved asparagine, histidine, and serine residues critical for catalysis. These residues form a charge relay network that facilitates the binding of substrate to protease. In addition to active site residues, other residues of the protease form an oxyanion hole. These residues stabilize
10 the substrate during catalysis and vary among different subclasses of serine proteases.

Other tissue-specific serine proteases have been discovered. In mouse, for example, the serine protease neuropsin is expressed primarily in the hippocampus, a region of the brain involved in long-term memory. Direct electrical stimulation of the hippocampus modulates neuropsin messenger RNA levels. Neuropsin cDNA predicts a
15 260-amino acid protein that shares 38% amino acid identity with pancreatic trypsin. Residues that form the catalytic triad and the oxyanion hole are conserved between the two proteins. A putative glycosylation site at asparagine 110 of neuropsin is contained within a unique consensus sequence, YNNSN, the same as that found in kallikrein-like serine proteases. Neuropsin may influence synapse formation and neuronal connectivity in the
20 hippocampus in response to neural signaling. (Chen, Z.-L. et al. (1995) J. Neurosci. 15:5088-5097.)

Prostate-specific antigen (PSA) is a serine protease synthesized and secreted exclusively by epithelial cells in the prostate gland. PSA is most related to kallikrein-like serine proteases. Inactive PSA is a 262-amino acid, 35-kilodalton glycoprotein that
25 contains a signal peptide sequence from amino acid residues 1-17 and an activation peptide from residues 18-24. Proteolytic removal of both these N-terminal peptides is required for activation of PSA. Ten cysteines in PSA form intramolecular disulfide bonds, and the catalytic triad is conserved. PSA is secreted into the seminal fluid in response to the male sex hormone, androgen. The substrate for PSA is seminogelin, a large protein
30 that promotes the coagulation of seminal fluid. Degradation of seminogelin by PSA results in the liquefaction of seminal fluid which allows increased sperm motility. (Henttu, P. and Vihko, P. (1994) Ann. Med. 26:157-164; and Riegman, P. H. J. et al.

(1989) Bioch. Biophys. Res. Commun. 159:95-102.)

Prostate cancer accounts for more than 13% of all cancer deaths and more than 28% of all cancers in men in the United States. These statistics are surpassed only by those reported for lung cancer. (McCance, K. L. and Huether, S. E. (1994)

5 Pathophysiology: The Biological Basis for Disease in Children and Adults, Mosby-Year Book, Inc., St. Louis, MO.) Normally, PSA is confined to the prostate, and only very low levels of PSA are detectable in serum. However, in the sera of patients with malignant prostate tumors, PSA levels are significantly elevated. The serum concentration of PSA generally increases with the severity of the clinical or pathological manifestations of the
10 disease. Currently, serum PSA is the most sensitive physiological marker for monitoring the progression of prostate cancer and its response to therapy. Serum PSA measurements in conjunction with other diagnostic tools may have application in early detection of the disease. PSA can also identify the prostate as the origin of a metastatic tumor. (Brawer, M. K. and Lange, P. H. (1989) Urology 33:11-16; and Henttu and Vihko, supra.)

15 Interestingly, the increase in serum PSA in prostate cancer patients is not due to increased expression of the gene encoding PSA. In fact, PSA gene expression is decreased as a result of disease. The accumulation of PSA in the serum may be attributed either to the disruption of the integrity of the prostate epithelium, which would result in the leakage of prostate contents, or to the depolarized secretion of PSA from the prostate epithelium
20 into areas accessible to the circulation. (Henttu and Vihko, supra.)

The discovery of a new prostate-associated serine protease and the polynucleotides encoding it satisfies a need in the art by providing new compositions which are useful in the diagnosis, treatment, and prevention of reproductive disorders and cancer.

25 SUMMARY OF THE INVENTION

The invention is based on the discovery of a human prostate-associated serine protease, PRASP, which shows homology to neuropsin, a brain-specific serine protease in mouse, and PSA, a prostate-specific serine protease in human. The invention features a substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or
30 a fragment of SEQ ID NO:1.

The invention further provides a substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of SEQ ID NO:1 or a fragment of

SEQ ID NO:1. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. The invention also includes an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention further provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, as well as an isolated and purified polynucleotide which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides an isolated and purified polynucleotide comprising the polynucleotide sequence of SEQ ID NO:2 or a fragment of SEQ ID NO:2, and an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide comprising the polynucleotide sequence of SEQ ID NO:2 or a fragment of SEQ ID NO:2. The invention also provides an isolated and purified polynucleotide having a sequence complementary to the polynucleotide comprising the polynucleotide sequence of SEQ ID NO:2 or a fragment of SEQ ID NO:2. The invention also provides polynucleotide fragments comprising nucleotides 146-181, 344-382, or 551-589 of SEQ ID NO:2, for detecting the presence or expression of an identical endogenous gene.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO: 1 under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the

host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1 in conjunction with a suitable pharmaceutical carrier.

5 The invention further includes a purified antibody which binds to a polypeptide comprising the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, as well as a purified agonist and a purified antagonist of the polypeptide.

The invention also provides a method for treating or preventing a reproductive disorder, the method comprising administering to a subject in need of such treatment an
10 effective amount of a pharmaceutical composition comprising substantially purified polypeptide having the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for treating or preventing a cancer, the method comprising administering to a subject in need of such treatment an effective
15 amount of a pharmaceutical composition comprising substantially purified polypeptide having the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for detecting a polynucleotide encoding a polypeptide comprising the amino acid sequence of SEQ ID NO: 1 or a fragment of SEQ ID NO: 1 in a biological sample containing nucleic acids, the method comprising the steps
20 of: (a) hybridizing the complement of the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1 to at least one of the nucleic acids of the biological sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide encoding the
25 polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1 in the biological sample. In one aspect, the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to the hybridizing step.

BRIEF DESCRIPTION OF THE FIGURES

30 Figures 1A, 1B, 1C, and 1D show the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ ID NO:2) of PRASP. The alignment was produced using MacDNASIS PRO™ software (Hitachi Software Engineering Co. Ltd., San Bruno, CA).

Figures 2A and 2B show the amino acid sequence alignments among PRASP (2723646; SEQ ID NO:1), mouse neuropsin (GI 1020091; SEQ ID NO:3), and human prostate-specific antigen (GI 511857; SEQ ID NO:4), produced using the multisequence alignment program of LASERGENE™ software (DNASTAR Inc, Madison WI).

5

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these may vary. It is also to be understood that
10 the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms “a,” “an,” and “the” include plural reference unless the context clearly dictates otherwise.
15 Thus, for example, a reference to “a host cell” includes a plurality of such host cells, and a reference to “an antibody” is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this
20 invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications and which might be used in
25 connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

“PRASP,” as used herein, refers to the amino acid sequences of substantially
30 purified PRASP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist," as used herein, refers to a molecule which, when bound to PRASP, increases or prolongs the duration of the effect of PRASP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of PRASP.

5 An "allele" or an "allelic sequence," as these terms are used herein, is an alternative form of the gene encoding PRASP. Alleles may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give
10 rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

 "Altered" nucleic acid sequences encoding PRASP, as described herein, include those sequences with deletions, insertions, or substitutions of different nucleotides,
15 resulting in a polynucleotide the same PRASP or a polypeptide with at least one functional characteristic of PRASP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PRASP, and improper or unexpected hybridization to alleles, with a locus other than the normal chromosomal locus for the polynucleotide sequence
20 encoding PRASP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PRASP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity
25 of PRASP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

30 The terms "amino acid" or "amino acid sequence," as used herein, refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments", "immunogenic

fragments", or "antigenic fragments" refer to fragments of PRASP which are preferably about 5 to about 15 amino acids in length and which retain some biological activity or immunological activity of PRASP. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid
5 sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification," as used herein, relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art. (See, e.g., Dieffenbach, C.W. and
10 G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, pp.1-5.)

The term "antagonist," as it is used herein, refers to a molecule which, when bound to PRASP, decreases the amount or the duration of the effect of the biological or immunological activity of PRASP. Antagonists may include proteins, nucleic acids,
15 carbohydrates, antibodies, or any other molecules which decrease the effect of PRASP.

As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fa, F(ab')₂, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind PRASP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the
20 immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the
25 animal.

The term "antigenic determinant," as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic
30 determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense," as used herein, refers to any composition containing a nucleic acid sequence which is complementary to a specific nucleic acid sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

As used herein, the term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic PRASP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity," as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition comprising a given amino acid sequence," as these terms are used herein, refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation, an aqueous solution, or a sterile composition. Compositions comprising polynucleotide sequences encoding PRASP or fragments of PRASP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS), and other components (e.g., Denhardt's solution, dry milk,

salmon sperm DNA, etc.).

The phrase "consensus sequence," as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using XL-PCR™ (Perkin Elmer, Norwalk, CT) in the 5' and/or the 3' direction, and resequenced, or which has been
5 assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW™ Fragment Assembly system (GCG, Madison, WI). Some sequences have been both extended and assembled to produce the consensus sequence .

As used herein, the term "correlates with expression of a polynucleotide" indicates
10 that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding PRASP, by northern analysis is indicative of the presence of nucleic acids encoding PRASP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding PRASP.

A "deletion," as the term is used herein, refers to a change in the amino acid or
15 nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative," as used herein, refers to the chemical modification of PRASP, of a polynucleotide sequence encoding PRASP, or of a polynucleotide sequence complementary to a polynucleotide sequence encoding PRASP. Chemical modifications of
20 a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains a at least one biological or immunological function of the polypeptide from which
25 it was derived.

The term "homology," as used herein, refers to a degree of complementarity. There may be partial homology or complete homology. The word "identity" may substitute for the word "homology." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred
30 to as "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of

reduced stringency. A substantially homologous sequence or hybridization probe will compete for and inhibit the binding of a completely homologous sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% homology or identity). In the absence of non-specific binding, the substantially homologous sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MegAlign™ program (DNASTAR, Inc., Madison WI). The MegAlign program can create alignments between two or more sequences according to different methods, e.g., the CLUSTAL V method. (See, e.g., Higgins, D.G. and P. M. Sharp (1988) *Gene* 73:237-244.) The CLUSTAL V algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no homology between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) *Methods Enzymol.* 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs), as described herein, are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance. (See, e.g., Harrington, J.J. et al. (1997) *Nat Genet.* 15:345-355.)

The term "humanized antibody," as used herein, refers to antibody molecules in

which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

“Hybridization,” as the term is used herein, refers to any process by which a strand
5 of nucleic acid binds with a complementary strand through base pairing.

As used herein, the term “hybridization complex” as used herein, refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present
10 in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words “insertion” or “addition,” as used herein, refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or
15 nucleotides, respectively, to the sequence found in the naturally occurring molecule.

“Immune response” can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

20 The term “microarray,” as used herein, refers to an arrangement of distinct polynucleotides arrayed on a substrate, e.g., paper, nylon or any other type of membrane, filter, chip, glass slide, or any other suitable solid support.

The terms “element” or “array element” as used herein in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

25 The term “modulate,” as it appears herein, refers to a change in the activity of PRASP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PRASP.

The phrases “nucleic acid” or “nucleic acid sequence,” as used herein, refer to an
30 oligonucleotide, nucleotide, polynucleotide, or any fragment thereof, to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-

like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which are greater than about 60 nucleotides in length, and most preferably are at least about 100 nucleotides, at least about 1000 nucleotides, or at least about 10,000 nucleotides in length.

5 The terms "operably associated" or "operably linked," as used herein, refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in reading frame, certain genetic elements, e.g., repressor genes, are not
10 contiguously linked to the encoded polypeptide but still bind to operator sequences that control expression of the polypeptide.

 The term "oligonucleotide," as used herein, refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in
15 a hybridization assay or microarray. As used herein, the term "oligonucleotide" is substantially equivalent to the terms "amplimer," "primer," "oligomer," and "probe," as these terms are commonly defined in the art.

 "Peptide nucleic acid" (PNA), as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in
20 length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA and RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell. (See, e.g., Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8:53-63.)

25 The term "sample," as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acids encoding PRASP, or fragments thereof, or PRASP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a solid support; a tissue; a tissue print; etc.

30 As used herein, the terms "specific binding" or "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the

antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

5 As used herein, the term "stringent conditions" refers to conditions which permit hybridization between polynucleotide sequences and the claimed polynucleotide sequences. Suitably stringent conditions can be defined by, for example, the concentrations of salt or formamide in the prehybridization and hybridization solutions, or by the hybridization temperature, and are well known in the art. In particular, stringency
10 can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

For example, hybridization under high stringency conditions could occur in about 50% formamide at about 37°C to 42°C. Hybridization could occur under reduced stringency conditions in about 35% to 25% formamide at about 30°C to 35°C. In
15 particular, hybridization could occur under high stringency conditions at 42°C in 50% formamide, 5X SSPE, 0.3% SDS, and 200 µg/ml sheared and denatured salmon sperm DNA. Hybridization could occur under reduced stringency conditions as described above, but in 35% formamide at a reduced temperature of 35°C. The temperature range corresponding to a particular level of stringency can be further narrowed by calculating the
20 purine to pyrimidine ratio of the nucleic acid of interest and adjusting the temperature accordingly. Variations on the above ranges and conditions are well known in the art.

The term "substantially purified," as used herein, refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably
25 about 90% free from other components with which they are naturally associated.

A "substitution," as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation," as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial
30 conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host

cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of PRASP, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

THE INVENTION

The invention is based on the discovery of a new human prostate-associated serine protease (PRASP), the polynucleotides encoding PRASP, and the use of these compositions for the diagnosis, treatment, or prevention of reproductive disorders and cancer.

Nucleic acids encoding the PRASP of the present invention were first identified in Incyte Clone 2723646 from the lung tumor cDNA library (LUNGTUT10) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 964204 (BRSTNOT05), 1685649 (PROSNOT15), and 2723646 (LUNGTUT10).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1, as shown in Figures 1A, 1B, 1C, and 1D. PRASP is 282 amino acids in length and has four potential N-glycosylation sites at N₁₃₁, N₁₉₇, N₂₁₃, and N₂₄₂; three potential casein kinase II phosphorylation sites at T₁₂₀, T₁₉₉, and S₂₂₂; five potential protein kinase C phosphorylation sites at S₁₃, S₁₆₄, T₁₉₂, T₂₅₉, and T₂₇₈; potential

signal peptide and activation peptide sequences from M₁ to G₅₀; and two serine protease trypsin family active site motifs from L₉₀ to C₉₅ and from D₂₂₉ to V₂₄₀. The region of PRASP from R₅₃ to I₂₇₅ also shows similarity to a domain of trypsin. As shown in Figures 2A and 2B, PRASP has chemical and structural homology with mouse neuropsin (GI 1020091; SEQ ID NO:3) and human PSA (GI 511857; SEQ ID NO:4). In particular, PRASP and mouse neuropsin share 47% identity, and PRASP and human PSA share 36% identity. In addition, residues of the active site catalytic triad are conserved in PRASP at H₉₄, D₁₄₂, and S₂₃₅. The residues that potentially form the oxyanion hole of neuropsin are conserved in PRASP at D₂₂₉, S₂₃₅, G₂₅₂, and G₂₆₃. The ten cysteines involved in intramolecular disulfide bond formation in PSA are conserved in neuropsin and in PRASP at C₆₀, C₇₉, C₉₅, C₁₇₄, C₁₉₅, C₂₀₆, C₂₂₀, C₂₃₁, C₂₄₁, and C₂₅₆. A potential N-glycosylation site in neuropsin is conserved in PRASP at N₁₃₁. Potential signal peptide sequences are conserved between PRASP and PSA. Fragments of SEQ ID NO:2 from about nucleotide 146 to about nucleotide 181; from about nucleotide 344 to about nucleotide 382; and from about nucleotide 551 to about nucleotide 589 are useful for distinguishing nucleotide sequences encoding PRASP from those encoding other known proteases. Northern analysis shows the expression of this sequence in various libraries, at least 81% of which are associated with cancer. In particular, 71% of the libraries expressing PRASP are derived from reproductive tissue, 57% from male reproductive tissue, and 48% specifically from prostate.

The invention also encompasses PRASP variants. A preferred PRASP variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the PRASP amino acid sequence, and which contains at least one functional or structural characteristic of PRASP.

The invention also encompasses polynucleotides which encode PRASP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:2, which encodes an PRASP.

The invention also encompasses a variant of a polynucleotide sequence encoding PRASP. In particular, such a variant polynucleotide sequence will have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PRASP. A particular aspect of the invention encompasses a variant of SEQ ID NO:2 which has at

least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to SEQ ID NO:2. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PRASP.

5 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PRASP, some bearing minimal homology to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based
10 on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PRASP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PRASP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring
15 PRASP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PRASP or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host.
20 Other reasons for substantially altering the nucleotide sequence encoding PRASP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode
25 PRASP and PRASP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PRASP or any fragment thereof.

30 Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:2, or a fragment of SEQ ID NO:2, under various conditions of stringency.

(See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; and Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.)

Methods for DNA sequencing are well known and generally available in the art and may be used to practice any of the embodiments of the invention. The methods may
5 employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp., Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE Amplification System (GIBCO/BRL, Gaithersburg, MD). Preferably, the process is automated with machines such as the
10 Hamilton Micro Lab 2200 (Hamilton, Reno, NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

The nucleic acid sequences encoding PRASP may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream
15 sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) In particular, genomic DNA is first amplified in the presence of a primer which is complementary to a linker sequence within the vector and a primer specific to a
20 region of the nucleotide sequence. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent
25 primers based on a known region. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) The primers may be designed using commercially available software such as OLIGO 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, MN) or another appropriate program to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of
30 about 68°C to 72°C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR, which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations
5 may be used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060.) Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This
10 process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable in that they will include more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T)
15 library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic
20 separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., Genotyper™ and Sequence Navigator™, Perkin Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer
25 controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PRASP may be used in recombinant DNA molecules to direct expression of PRASP, or fragments or functional equivalents thereof, in appropriate host
30 cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced, and these sequences may be used to clone and express PRASP.

As will be understood by those of skill in the art, it may be advantageous to produce PRASP-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce an RNA transcript
5 having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PRASP-encoding sequences for a variety of reasons including, but not limited to, alterations which modify the cloning,
10 processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

15 In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PRASP may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of PRASP activity, it may be useful to encode a chimeric PRASP protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a
20 cleavage site located between the PRASP encoding sequence and the heterologous protein sequence, so that PRASP may be cleaved and purified away from the heterologous moiety.

In another embodiment, sequences encoding PRASP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232.) Alternatively, the protein itself may be produced using chemical
25 methods to synthesize the amino acid sequence of PRASP, or a fragment thereof. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A Peptide Synthesizer (Perkin Elmer). Additionally, the
30 amino acid sequence of PRASP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g, Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1983) Proteins, Structures and
5 Molecular Properties, WH Freeman and Co., New York, NY.)

In order to express a biologically active PRASP, the nucleotide sequences encoding PRASP or derivatives thereof may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

10 Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PRASP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press,
15 Plainview, NY, ch. 4, 8, and 16-17; and Ausubel, F.M. et al. (1995, and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PRASP. These include, but are not limited to, microorganisms such
20 as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV)) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or
25 animal cell systems.

The invention is not limited by the host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions, e.g., enhancers, promoters, and 5' and 3' untranslated regions, of the vector and polynucleotide sequences encoding PRASP which interact with host cellular proteins to
30 carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may

be used. For example, when cloning in bacterial systems, inducible promoters, e.g., hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, La Jolla, CA) or pSport1™ plasmid (GIBCO/BRL), may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO, and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding PRASP, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for PRASP. For example, when large quantities of PRASP are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, multifunctional *E. coli* cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding PRASP may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of β -galactosidase so that a hybrid protein is produced, and pIN vectors. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) pGEX vectors (Pharmacia Biotech, Uppsala, Sweden) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used. (See, e.g., Ausubel, *supra*; and Grant et al. (1987) Methods Enzymol. 153:516-544.)

In cases where plant expression vectors are used, the expression of sequences encoding PRASP may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) EMBO

J. 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells
5 by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews. (See, e.g., Hobbs, S. or Murry, L.E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.)

An insect system may also be used to express PRASP. For example, in one such
10 system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding PRASP may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of sequences encoding PRASP will render the polyhedrin gene inactive and
15 produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells or Trichoplusia larvae in which PRASP may be expressed. (See, e.g., Engelhard, E.K. et al. (1994) Proc. Nat. Acad. Sci. 91:3224-3227.)

In mammalian host cells, a number of viral-based expression systems may be
20 utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PRASP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing PRASP in infected host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc.
25 Natl. Acad. Sci. 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of about 6 kb
30 to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation of

sequences encoding PRASP. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding PRASP and its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where
5 only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of
10 enhancers appropriate for the particular cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation,
15 glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding, and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture
20 Collection (ATCC, Bethesda, MD) and may be chosen to ensure the correct modification and processing of the foreign protein.

For long term, high yield production of recombinant proteins, stable expression is preferred. For example, cell lines capable of stably expressing PRASP can be transformed using expression vectors which may contain viral origins of replication and/or endogenous
25 expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of
30 stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines.

These include, but are not limited to, the herpes simplex virus thymidine kinase genes and adenine phosphoribosyltransferase genes, which can be employed in *tk* or *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; and Lowy, I. et al. (1980) Cell 22:817-823) Also, antimetabolite, antibiotic, or herbicide resistance can be
5 used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *npt* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14; and Murry, supra.) Additional selectable genes have been
10 described, e.g., *trpB*, which allows cells to utilize indole in place of tryptophan, or *hisD*, which allows cells to utilize histinol in place of histidine. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Visible markers, e.g., anthocyanins, β glucuronidase and its substrate GUS, luciferase and its substrate luciferin may be used. Green fluorescent proteins (GFP) (Clontech, Palo Alto, CA) can also be
15 used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed.
20 For example, if the sequence encoding PRASP is inserted within a marker gene sequence, transformed cells containing sequences encoding PRASP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PRASP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the
25 tandem gene as well.

Alternatively, host cells which contain the nucleic acid sequence encoding PRASP and express PRASP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane,
30 solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

The presence of polynucleotide sequences encoding PRASP can be detected by

DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding PRASP. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding PRASP to detect transformants containing DNA or RNA encoding PRASP.

5 A variety of protocols for detecting and measuring the expression of PRASP, using either polyclonal or monoclonal antibodies specific for the protein, are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two
10 non-interfering epitopes on PRASP is preferred, but a competitive binding assay may be employed. These and other assays are well described in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Section IV; and Maddox, D.E. et al. (1983) J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in
15 the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PRASP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PRASP, or any fragments thereof, may be cloned into a vector for the production of an
20 mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Pharmacia & Upjohn (Kalamazoo, MI), Promega (Madison, WI), and U.S. Biochemical Corp. (Cleveland, OH).
25 Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PRASP may be cultured under conditions suitable for the expression and recovery of the protein from cell
30 culture. The protein produced by a transformed cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode

PRASP may be designed to contain signal sequences which direct secretion of PRASP through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding PRASP to nucleotide sequences encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences, such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA), between the purification domain and the PRASP encoding sequence may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing PRASP and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on immobilized metal ion affinity chromatography (IMAC). (See, e.g., Porath, J. et al. (1992) Prot. Exp. Purif. 3: 263-281.) The enterokinase cleavage site provides a means for purifying PRASP from the fusion protein. (See, e.g., Kroll, D.J. et al. (1993) DNA Cell Biol. 12:441-453.)

Fragments of PRASP may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, T.E. (1984) Protein: Structures and Molecular Properties, pp. 55-60, W.H. Freeman and Co., New York, NY.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of PRASP may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural homology exists among PRASP, neuropsin from mouse (GI 1020091), and PSA from human (GI 511857). In addition, PRASP is expressed in reproductive tissue, particularly in the prostate. Therefore, PRASP appears to play a role in reproductive disorders and cancer.

Therefore, in one embodiment, PRASP or a fragment or derivative thereof may be administered to a subject to treat or prevent a reproductive disorder. Such disorders can

include, but are not limited to, abnormal prolactin production, infertility, tubal disease, ovulatory defects, endometriosis, perturbations of the estrous and menstrual cycles, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, autoimmune disorders, ectopic pregnancy, teratogenesis, breast cancer, fibrocystic breast disease, galactorrhea, abnormal spermatogenesis, abnormal sperm physiology, testicular cancer, prostate cancer, benign prostatic hyperplasia, prostatitis, and gynecomastia.

In another embodiment, a vector capable of expressing PRASP or a fragment or derivative thereof may be administered to a subject to treat or prevent a reproductive disorder including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified PRASP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a reproductive disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PRASP may be administered to a subject to treat or prevent a reproductive disorder including, but not limited to, those listed above.

In another embodiment, PRASP or a fragment or derivative thereof may be administered to a subject to treat or prevent a cancer. Such cancers can include, but are not limited to, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing PRASP or a fragment or derivative thereof may be administered to a subject to treat or prevent a cancer including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified PRASP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a cancer including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PRASP may be administered to a subject to treat or prevent a cancer including, but not limited to,

those listed above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PRASP may be produced using methods which are generally known in the art. In particular, purified PRASP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PRASP. Antibodies to PRASP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PRASP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PRASP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of PRASP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be

produced.

Monoclonal antibodies to PRASP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture.

These include, but are not limited to, the hybridoma technique, the human B-cell

- 5 hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci.* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

- 10 In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci.* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.)

- 15 Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PRASP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) *Proc. Natl. Acad. Sci.* 88:10134-10137.)

- 20 Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci.* 86: 3833-3837; and Winter, G. et al. (1991) *Nature* 349:293-299.)

- 25 Antibody fragments which contain specific binding sites for PRASP may also be generated. For example, such fragments include, but are not limited to, F(ab')₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

- 30 Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are

well known in the art. Such immunoassays typically involve the measurement of complex formation between PRASP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PRASP epitopes is preferred, but a competitive binding assay may also be employed. (Maddox, supra.)

In another embodiment of the invention, the polynucleotides encoding PRASP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding PRASP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding PRASP. Thus, complementary molecules or fragments may be used to modulate PRASP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PRASP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors which will express nucleic acid sequences complementary to the polynucleotides of the gene encoding PRASP. (See, e.g., Sambrook, supra; and Ausubel, supra.)

Genes encoding PRASP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding PRASP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding PRASP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the

start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have
5 been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific
10 cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PRASP.

15 Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may
20 render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules.
25 These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PRASP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that
30 synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life.

Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) *Nature Biotechnology* 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of PRASP, antibodies to PRASP, and mimetics, agonists, antagonists, or inhibitors of PRASP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used

pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

Pharmaceutical compositions for oral administration can be formulated using
5 pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active
10 compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose,
15 hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as
20 concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

25 Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or
30 suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be

formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the
5 active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the
10 preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in
15 a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric,
20 malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

25 After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of PRASP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions
30 wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to
5 determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PRASP or fragments thereof, antibodies of PRASP, and agonists, antagonists or inhibitors of PRASP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or
10 with experimental animals, such as by calculating the ED₅₀ (the dose therapeutically effective in 50% of the population) or LD₅₀ (the dose lethal to 50% of the population) statistics. The dose ratio of therapeutic to toxic effects is the therapeutic index, and it can be expressed as the ED₅₀/LD₅₀ ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal
15 studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED₅₀ with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related
20 to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting
25 pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μ g to 100,000 μ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally
30 available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations,

etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind PRASP may be used
5 for the diagnosis of disorders characterized by expression of PRASP, or in assays to
monitor patients being treated with PRASP or agonists, antagonists, or inhibitors of
PRASP. Antibodies useful for diagnostic purposes may be prepared in the same manner
as described above for therapeutics. Diagnostic assays for PRASP include methods which
utilize the antibody and a label to detect PRASP in human body fluids or in extracts of
10 cells or tissues. The antibodies may be used with or without modification, and may be
labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of
reporter molecules, several of which are described above, are known in the art and may be
used.

A variety of protocols for measuring PRASP, including ELISAs, RIAs, and FACS,
15 are known in the art and provide a basis for diagnosing altered or abnormal levels of
PRASP expression. Normal or standard values for PRASP expression are established by
combining body fluids or cell extracts taken from normal mammalian subjects, preferably
human, with antibody to PRASP under conditions suitable for complex formation. The
amount of standard complex formation may be quantitated by various methods, preferably
20 by photometric means. Quantities of PRASP expressed in subject, control, and disease
samples from biopsied tissues are compared with the standard values. Deviation between
standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PRASP may
be used for diagnostic purposes. The polynucleotides which may be used include
25 oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The
polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in
which expression of PRASP may be correlated with disease. The diagnostic assay may be
used to determine absence, presence, and excess expression of PRASP, and to monitor
regulation of PRASP levels during therapeutic intervention.

30 In one aspect, hybridization with PCR probes which are capable of detecting
polynucleotide sequences, including genomic sequences, encoding PRASP or closely
related molecules may be used to identify nucleic acid sequences which encode PRASP.

The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding

5 PRASP, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the PRASP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:2 or from genomic sequences including

10 promoters, enhancers, and introns of the PRASP gene.

Means for producing specific hybridization probes for DNAs encoding PRASP include the cloning of polynucleotide sequences encoding PRASP or PRASP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of

15 the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PRASP may be used for the diagnosis of a

20 disorder associated with expression of PRASP. Examples of such a disorder include, but are not limited to, reproductive disorders, such as abnormal prolactin production, infertility, tubal disease, ovulatory defects, endometriosis, perturbations of the estrous and menstrual cycles, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, autoimmune disorders, ectopic pregnancy, teratogenesis,

25 breast cancer, fibrocystic breast disease, galactorrhea, abnormal spermatogenesis, abnormal sperm physiology, testicular cancer, prostate cancer, benign prostatic hyperplasia, prostatitis, carcinoma of the male breast, and gynecomastia; and cancers, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone

30 marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding

PRASP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and ELISA assays; and in microarrays utilizing fluids or tissues from patients to detect altered PRASP expression. Such qualitative or quantitative methods are well known in the art.

5 In a particular aspect, the nucleotide sequences encoding PRASP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PRASP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and
10 the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PRASP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in
15 clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PRASP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PRASP, under
20 conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard
25 values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of
30 treatment over a period ranging from several days to months.

With respect to prostate cancer, the presence of a relatively high amount of transcript in serum from an individual may indicate a predisposition for the development

of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

5 Additional diagnostic uses for oligonucleotides designed from the sequences encoding PRASP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PRASP, or a fragment of a polynucleotide complementary to the polynucleotide encoding PRASP, and will be
10 employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of PRASP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and
15 interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; and Duplaa, C. et al. (1993) Anal. Biochem. 229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

20 In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a
25 disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application
30 WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding PRASP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human
5 artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Price, C.M. (1993) *Blood Rev.* 7:127-134; and Trask, B.J. (1991) *Trends Genet.* 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical
10 chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, R.A. (ed.) Molecular Biology and Biotechnology, VCH Publishers New York, NY, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding PRASP on a physical chromosomal map and a specific
15 disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used
20 for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery
25 techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., AT to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation,
30 inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PRASP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in

any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PRASP and the agent being tested may be measured.

5 Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with PRASP, or fragments thereof, and
10 washed. Bound PRASP is then detected by methods well known in the art. Purified PRASP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which
15 neutralizing antibodies capable of binding PRASP specifically compete with a test compound for binding PRASP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PRASP.

In additional embodiments, the nucleotide sequences which encode PRASP may be used in any molecular biology techniques that have yet to be developed, provided the new
20 techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

25

EXAMPLES

I. LUNGTUT10 cDNA Library Construction

The LUNGTUT10 cDNA library was constructed from lung tumor tissue removed from a 65 year-old Caucasian female during a segmental lung resection. Pathology of the
30 left upper lung indicated a metastatic grade 2 myxoid liposarcoma. In addition, pathology of a tumor mass above the diaphragm and between the lungs indicated a metastatic grade 4 liposarcoma. Patient history included malignant soft tissue neoplasm, malignant breast

neoplasm, and secondary lung neoplasm.

The frozen tissue was homogenized and lysed in TRIzol reagent (1 gm tissue/10 ml TRIzol; Catalog #10296-028, Gibco/BRL, Gaithersburg, MD), a monoplastic solution of phenol and guanidine isothiocyanate, using a Brinkmann Homogenizer Polytron PT-3000
5 (Brinkmann Instruments, Westbury, NY). After brief incubation on ice, chloroform was added (1:5 v/v), and the mixture was centrifuged to separate the phases. The upper aqueous phase was removed to a fresh tube, and isopropanol was added to precipitate RNA. The RNA was resuspended in RNase-free water and treated with DNase. The RNA was re-extracted twice with acid phenol-chloroform and reprecipitated with sodium acetate
10 and ethanol. Poly(A+) RNA was isolated using the Qiagen Oligotex kit (QIAGEN, Chatsworth, CA).

Poly (A+) RNA was used to construct the LUNGTUT10 cDNA library according to the recommended protocols in the SuperScript Plasmid System (Catalog #18248-013, Gibco/BRL). The cDNAs were fractionated on a Sepharose CL4B column (Catalog
15 #275105-01, Pharmacia, Piscataway, NJ), and those cDNAs exceeding 400 bp were ligated into the plasmid pINCY 1 (Incyte), a derivative of pSPORT I (Gibco/BRL). pINCY 1 was subsequently transformed into DH5 α TM competent cells (Catalog #18258-012, Gibco/BRL).

20 II. Isolation and Sequencing of cDNA Clones

Plasmid DNA was released from the cells and purified using the REAL Prep 96 plasmid kit (Catalog #26173, QIAGEN). The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Catalog #22711, Gibco/BRL) with carbenicillin at 25 mg/L and glycerol at 0.4%;
25 2) after the cultures were incubated for 19 hours, the cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellets were each resuspended in 0.1 ml of distilled water. The DNA samples were stored at 4° C.

The cDNAs were sequenced by the method of Sanger et al. (1975, J. Mol. Biol. 94:441f), using a Hamilton Micro Lab 2200 (Hamilton, Reno, NV) in combination with
30 Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown, MA) and Applied Biosystems 377 DNA Sequencing Systems.

III. Homology Searching of cDNA Clones and Their Deduced Proteins

The nucleotide sequences and/or amino acid sequences of the Sequence Listing were used to query sequences in the GenBank, SwissProt, BLOCKS, and Pima II
5 databases. These databases, which contain previously identified and annotated sequences, were searched for regions of homology using BLAST (Basic Local Alignment Search Tool). (See, e.g., Altschul, S.F. (1993) J. Mol. Evol 36:290-300; and Altschul et al. (1990) J. Mol. Biol. 215:403-410.)

BLAST produced alignments of both nucleotide and amino acid sequences to
10 determine sequence similarity. Because of the local nature of the alignments, BLAST was especially useful in determining exact matches or in identifying homologs which may be of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. Other algorithms could have been used when dealing with primary sequence patterns and secondary structure gap penalties. (See, e.g., Smith, T. et al. (1992) Protein Engineering 5:35-51.)
15 The sequences disclosed in this application have lengths of at least 49 nucleotides and have no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

The BLAST approach searched for matches between a query sequence and a database sequence. BLAST evaluated the statistical significance of any matches found, and reported only those matches that satisfy the user-selected threshold of significance. In
20 this application, threshold was set at 10^{-25} for nucleotides and 10^{-8} for peptides.

Incyte nucleotide sequences were searched against the GenBank databases for primate (pri), rodent (rod), and other mammalian sequences (mam), and deduced amino acid sequences from the same clones were then searched against GenBank functional protein databases, mammalian (mamp), vertebrate (vrtp), and eukaryote (eukp), for
25 homology.

Additionally, sequences identified from cDNA libraries may be analyzed to identify those gene sequences encoding conserved protein motifs using an appropriate analysis program, e.g., the Block 2 Bioanalysis Program (Incyte, Palo Alto, CA). This motif analysis program, based on sequence information contained in the Swiss-Prot
30 Database and PROSITE, is a method of determining the function of uncharacterized proteins translated from genomic or cDNA sequences. (See, e.g., Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; and Attwood, T. K. et al. (1997) J. Chem. Inf. Comput.

Sci. 37:417-424.) PROSITE may be used to identify common functional or structural domains in divergent proteins. The method is based on weight matrices. Motifs identified by this method are then calibrated against the SWISS-PROT database in order to obtain a measure of the chance distribution of the matches.

- 5 In another alternative, Hidden Markov models (HMMs) may be used to find protein domains, each defined by a dataset of proteins known to have a common biological function. (See, e.g., Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; and Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197.) HMMs were initially developed to examine speech recognition patterns, but are now being
- 10 used in a biological context to analyze protein and nucleic acid sequences as well as to model protein structure. (See, e.g., Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; and Collin, M. et al. (1993) Protein Sci. 2:305-314.) HMMs have a formal probabilistic basis and use position-specific scores for amino acids or nucleotides. The algorithm continues to incorporate information from newly identified sequences to increase its motif
- 15 analysis capabilities.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a

20 membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; and Ausubel, F.M. et al. *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST are used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ™ database (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based

25 hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

30

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match

will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analysis are reported as a list of libraries in which the
5 transcript encoding PRASP occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

10 V. Extension of PRASP Encoding Polynucleotides

The nucleic acid sequence of Incyte Clone 2723646 was used to design oligonucleotide primers for extending a partial nucleotide sequence to full length. One primer was synthesized to initiate extension of an antisense polynucleotide, and the other was synthesized to initiate extension of a sense polynucleotide. Primers were used to
15 facilitate the extension of the known sequence "outward" generating amplicons containing new unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO 4.06 (National Biosciences, Plymouth, MN), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of
20 about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (GIBCO/BRL) were used to extend the sequence. If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

25 High fidelity amplification was obtained by following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix. PCR was performed using the Peltier Thermal Cycler (PTC200; M.J. Research, Watertown, MA), beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, with the following parameters:

30	Step 1	94° C for 1 min (initial denaturation)
	Step 2	65° C for 1 min
	Step 3	68° C for 6 min
	Step 4	94° C for 15 sec

- | | | |
|---|---------|---|
| | Step 5 | 65° C for 1 min |
| | Step 6 | 68° C for 7 min |
| | Step 7 | Repeat steps 4 through 6 for an additional 15 cycles |
| | Step 8 | 94° C for 15 sec |
| 5 | Step 9 | 65° C for 1 min |
| | Step 10 | 68° C for 7:15 min |
| | Step 11 | Repeat steps 8 through 10 for an additional 12 cycles |
| | Step 12 | 72° C for 8 min |
| | Step 13 | 4° C (and holding) |

10

A 5 μ l to 10 μ l aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6% to 0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQuick™ (QIAGEN Inc., Chatsworth, CA), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13 μ l of ligation buffer, 1 μ l T4-DNA ligase (15 units) and 1 μ l T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2 to 3 hours, or overnight at 16° C. Competent *E. coli* cells (in 40 μ l of appropriate media) were transformed with 3 μ l of ligation mixture and cultured in 80 μ l of SOC medium. (See, e.g., Sambrook, *supra*, Appendix A, p. 2.) After incubation for one hour at 37° C, the *E. coli* mixture was plated on Luria Bertani (LB) agar (See, e.g., Sambrook, *supra*, Appendix A, p. 1) containing 2x Carb. The following day, several colonies were randomly picked from each plate and cultured in 150 μ l of liquid LB/2x Carb medium placed in an individual well of an appropriate commercially-available sterile 96-well microtiter plate. The following day, 5 μ l of each overnight culture was transferred into a non-sterile 96-well plate and, after dilution 1:10 with water, 5 μ l from each sample was transferred into a PCR array.

For PCR amplification, 18 μ l of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

- | | | |
|----|--------|--|
| | Step 1 | 94° C for 60 sec |
| | Step 2 | 94° C for 20 sec |
| | Step 3 | 55° C for 30 sec |
| 35 | Step 4 | 72° C for 90 sec |
| | Step 5 | Repeat steps 2 through 4 for an additional 29 cycles |
| | Step 6 | 72° C for 180 sec |

Step 7 4° C (and holding)

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequence of SEQ ID NO:2 is used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

10 **VI. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:2 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software
15 such as OLIGO 4.06 (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ -³²P] adenosine triphosphate (Amersham, Chicago, IL), and T4 polynucleotide kinase (DuPont NEN®, Boston, MA). The labeled oligonucleotides are substantially purified using a Sephadex G-25 superfine resin column (Pharmacia & Upjohn, Kalamazoo, MI). An aliquot containing 10⁷ counts per minute of the labeled
20 probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN, Boston, MA).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, NH).
25 Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak, Rochester, NY) is exposed to the blots to film for several hours, hybridization patterns are compared visually.

30

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize

array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any
5 appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may
10 comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE™. Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed
15 to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; and Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

20

VIII. Complementary Polynucleotides

Sequences complementary to the PRASP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PRASP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described,
25 essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using Oligo 4.06 software and the coding sequence of PRASP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to
30 prevent ribosomal binding to the PRASP-encoding transcript.

IX. Expression of PRASP

Expression of PRASP is accomplished by subcloning the cDNA into an appropriate vector and transforming the vector into host cells. This vector contains an appropriate promoter, e.g., β -galactosidase, upstream of the cloning site, operably associated with the cDNA of interest. (See, e.g., Sambrook, *supra*, pp. 404-433; and

5 Rosenberg, M. et al. (1983) *Methods Enzymol.* 101:123-138.)

Induction of an isolated, transformed bacterial strain with isopropyl beta-D-thiogalactopyranoside (IPTG) using standard methods produces a fusion protein which consists of the first 8 residues of β -galactosidase, about 5 to 15 residues of linker, and the full length protein. The signal residues direct the secretion of PRASP into bacterial
10 growth media which can be used directly in the following assay for activity.

X. Demonstration of PRASP Activity

An assay for PRASP activity measures the hydrolysis of a synthetic peptide that serves as a universal substrate for chymotrypsin-like serine proteases, including PSA.

15 (Christensson, A. et al. (1990) 194:755-763.) This peptide is synthesized by methods well known in the art. Its sequence, which is comprised of amino acids and organic functional groups, is represented as methyl-O-succinyl-arginine-proline-tyrosine-NH-*p*-nitroanilide. Approximately 30 μ g of PRASP, either biochemically purified from prostate or produced by recombinant methods, are added to 1 mM peptide in 1 ml buffered solution, pH 7.8.
20 Hydrolysis of the peptide substrate is measured spectrophotometrically at 405 nm wavelength. Kinetics of serine protease activity are characterized by a burst phase of rapid hydrolysis for about 2 minutes, followed by a linear rate of hydrolysis for several minutes. PRASP activity is demonstrated by its unique kinetics and is proportional to the amount of hydrolyzed peptide.

25

XI. Production of PRASP Specific Antibodies

PRASP substantially purified using PAGE electrophoresis (see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols. The PRASP amino
30 acid sequence is analyzed using DNASTAR software (DNASTAR Inc) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of

appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel et al. supra, ch. 11.)

Typically, the oligopeptides are 15 residues in length, and are synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry and
5 coupled to KLH (Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel et al. supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and
10 reacting with radio-iodinated goat anti-rabbit IgG.

XII. Purification of Naturally Occurring PRASP Using Specific Antibodies

Naturally occurring or recombinant PRASP is substantially purified by immunoaffinity chromatography using antibodies specific for PRASP. An
15 immunoaffinity column is constructed by covalently coupling anti-PRASP antibody to an activated chromatographic resin, such as CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PRASP are passed over the immunoaffinity column, and the
20 column is washed under conditions that allow the preferential absorbance of PRASP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PRASP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PRASP is collected.

25 XIII. Identification of Molecules Which Interact with PRASP

PRASP, or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PRASP, washed, and any wells with labeled PRASP complex are assayed. Data
30 obtained using different concentrations of PRASP are used to calculate values for the number, affinity, and association of PRASP with the candidate molecules.

Various modifications and variations of the described methods and systems of the

invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications
5 of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

What is claimed is:

1. A substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.
- 5 2. A substantially purified variant having at least 90% amino acid sequence identity to the sequence of claim 1.
3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.
- 10 4. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 3.
- 15 5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.
6. An isolated and purified polynucleotide which is complementary to the polynucleotide of claim 3.
- 20 7. An isolated and purified polynucleotide comprising the polynucleotide sequence of SEQ ID NO:2 or a fragment of SEQ ID NO:2.
8. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 7.
- 25 9. An isolated and purified polynucleotide having a sequence complementary to the polynucleotide of claim 7.
- 30 10. A fragment of the polynucleotide of claim 7 comprising nucleotides 146-181, 344-382, or 551-589 of SEQ ID NO:2.

11. An expression vector containing at least a fragment of the polynucleotide of claim 3.

12. A host cell containing the expression vector of claim 11.

5

13. A method for producing a polypeptide comprising a sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, the method comprising the steps of:

(a) culturing the host cell of claim 12 under conditions suitable for the expression of the polypeptide; and

10 (b) recovering the polypeptide from the host cell culture.

14. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.

15 15. A purified antibody which specifically binds to the polypeptide of claim 1.

16. A purified agonist of the polypeptide of claim 1.

17. A purified antagonist of the polypeptide of claim 1.

20

18. A method for treating or preventing a reproductive disorder, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 14.

25 19. A method for treating or preventing a cancer, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 14.

20. A method for detecting a polynucleotide encoding a polypeptide
30 comprising the amino acid sequence of SEQ ID NO:1 in a biological sample containing nucleic acids, the method comprising the steps of:

(a) hybridizing the polynucleotide of claim 6 to at least one of the

nucleic acids of the biological sample, thereby forming a hybridization complex;
and

- 5 (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide encoding the polypeptide in the biological sample.

21. The method of claim 20 wherein the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to the hybridizing step.

10	19	28	37	46	55
AACC TAG GCC CTC CCC TGC CTT GCT CCA CAC CTG GTC AGG GGA GAG ACG GGA GGA					
64	73	82	91	100	109
AAG CCA AGG GAA GGG ACC TAA CTG AAA ACA AAC AAG CTG GGA GAA GCA GGA ATC					
118	127	136	145	154	163
TGC GCT CGG GTT CCG CAG ATG CAG AGG TTG AGG TGG CTG CGG GAC TGG AAG TCA					
	M	Q	R	L	W
					K
					S
172	181	190	199	208	217
TCG GGC AGA GGT CTC ACA GCA GCC AAG GAA CCT GGG GCC CGC TCC TCC CCC CTC					
S	G	R	G	L	T
					A
					K
					E
					P
					G
					L
226	235	244	253	262	271
CAG GCC ATG AGG ATT CTG CAG TTA ATC CTG CTT GCT CTG GCA ACA GGG CTT GTA					
Q	A	M	R	I	L
					Q
					L
					A
					T
					G
					L
					V
280	289	298	307	316	325
GGG GGA GAG ACC AGG ATC ATC AAG GGG TTC GAG TGC TGC AAG CCT CAC TCC CAG CCC					
G	G	E	T	R	I
					A
					F
					C
					K
					P
					H
					S
					Q
					P
334	343	352	361	370	379
TGG CAG GCA GCC CTG TTC GAG AAG ACG CGG CTA CTC TGT GGG GCG ACG CTC ATC					
W	Q	A	A	L	F
					E
					K
					T
					R
					L
					L
					C
					G
					A
					T
					L
					I

FIGURE 1A

FIGURE 1B

FIGURE 1B

```

766   775   784   793   802   811
CCC GGC AAC ATC ACA GAC ACC ATG GTG TGT GCC AGC GTG CAG GAA GGC GGC AAG
P   G   N   I   T   D   T   M   V   C   A   S   V   Q   E   G   G   K

820   829   838   847   856   865
GAC TCC TGC CAG GGT GAC TCC GGC CCT CTG GTC TGT AAC CAG TCT CTT CAA
D   S   C   Q   G   D   S   G   G   P   L   V   C   N   Q   S   L   Q

874   883   892   901   910   919
GGC ATT ATC TCC TGG GGC CAG GAT CCG TGT GCG ATC ACC CGA AAG CCT GGT GTC
G   I   I   S   W   G   G   Q   D   P   C   A   I   T   R   K   P   G   V

928   937   946   955   964   973
TAC ACG AAA GTC TGC AAA TAT GTG GAC TGG ATC CAG GAG ACG ATG AAG AAC AAT
Y   T   K   V   C   K   Y   V   D   W   I   Q   E   T   M   K   N   N

982   991   1000   1009   1018   1027
TAG ACT GGA CCC ACC CAC CAC AGC CCA TCA CCC TCC ATT TCC ACT TGG TGT TTG

1036   1045   1054   1063   1072   1081
GTT CCT GTT CAC TCT GTT AAT AAG AAA CCC TAA GCC AAG ACC CTC TAC GAA CAT

1090   1099   1108   1117   1126   1135
TCT TTG GGC CTC CTG GAC TAC AGG AGA TGC TGT CAC TTA ATA ATC AAC CTG GGG

1144   1153   1162   1171   1180   1189
TTC GAA ATC AGT GAG ACC TGG ATT CAA ATT CTG CCT TGA AAT ATT GTG ACT CTG

```

FIGURE 1C

1198 GGA ATG ACA ACA CCT GGT TTG TTC TCT GTT GTA TCC CCA GCC CCA AAG ACA GCT 1243
1207
1216
1225
1234
1243
1252 CCT GGC CAT ATA TCA AGG TTT CAA TAA ATA TTT GCT AAA TGA AAA AAA AAA AAA 1297
1261
1270
1279
1288
1297
1306 AAA GGG CGG TCG CTC TA

FIGURE 1D

1	M	Q	R	L	R	W	L	R	D	W	K	S	S	G	R	G	L	T	A	A	K	E	P	G	A	R	S	S	P	L	2723646
1	M	G	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P	P	P	C	A	I	GI 1020091
1	M	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 511857
31	Q	A	M	R	I	L	Q	L	I	L	A	L	A	T	G	L	V	G	G	E	-	T	R	I	I	K	G	F	E	2723646	
10	Q	P	-	W	I	L	L	L	L	F	M	G	A	W	A	G	L	T	R	A	Q	G	S	K	I	L	E	G	R	E	GI 1020091
5	-	-	-	-	V	V	F	L	T	L	S	V	T	W	I	G	A	P	L	I	L	S	R	I	V	G	G	W	E	GI 511857	
60	C	K	P	H	S	Q	P	W	Q	A	A	L	F	E	K	T	R	L	L	C	G	A	T	L	I	A	P	R	W	L	2723646
39	C	I	P	H	S	Q	P	W	Q	A	A	L	F	Q	G	E	R	L	I	C	G	G	V	L	V	G	D	R	W	V	GI 1020091
31	C	E	K	H	S	Q	P	W	Q	V	L	V	A	S	R	G	R	A	V	C	G	G	V	L	V	H	P	Q	W	V	GI 511857
90	L	T	A	A	H	C	L	K	P	-	R	Y	I	V	H	L	G	Q	H	N	L	Q	K	E	E	G	C	E	Q	T	2723646
69	L	T	A	A	H	C	K	K	Q	-	K	Y	S	V	R	L	G	D	H	S	L	Q	S	R	D	Q	P	E	Q	E	GI 1020091
61	L	T	A	A	H	C	I	R	K	C	K	S	V	I	L	L	G	R	H	S	L	F	H	P	E	D	T	G	Q	V	GI 511857
119	R	T	A	T	E	S	F	P	H	P	G	F	N	N	S	L	-	-	-	-	-	-	-	P	N	K	D	H	R	N	2723646
98	I	Q	V	A	Q	S	I	Q	H	P	C	Y	N	N	S	N	-	-	-	-	-	-	-	P	E	-	D	H	S	H	GI 1020091
91	F	Q	V	S	H	S	F	P	H	P	L	Y	D	M	S	L	L	K	N	R	F	L	R	P	G	D	D	S	S	H	GI 511857

FIGURE 2A

142	D	I	M	L	V	K	M	A	S	P	V	S	I	T	W	A	V	R	P	L	T	L	S	S	R	C	V	T	A	G	2723646	
120	D	I	M	L	I	R	L	Q	N	S	A	N	L	G	D	K	V	K	P	V	Q	L	A	N	L	C	P	K	V	G	GI 1020091	
121	D	L	M	L	L	R	L	S	E	P	A	E	L	T	D	A	V	K	V	M	D	L	P	T	Q	E	P	A	L	G	GI 511857	
172	T	S	C	L	I	S	G	W	G	S	T	S	S	P	Q	L	R	L	P	H	T	L	R	C	A	N	I	T	I	I	2723646	
150	Q	K	C	I	I	S	G	W	G	T	V	T	S	S	P	Q	E	N	F	P	N	T	L	N	C	A	E	V	K	I	Y	GI 1020091
151	T	T	C	Y	A	S	G	W	G	S	I	E	P	E	E	F	L	T	P	K	K	L	Q	C	V	D	L	H	V	I	GI 511857	
202	E	H	Q	K	C	E	N	A	Y	P	G	N	I	T	D	T	M	V	C	A	S	V	Q	E	G	G	K	D	S	C	2723646	
180	S	Q	N	K	C	E	R	A	Y	P	G	K	I	T	E	G	M	V	C	A	G	S	S	N	G	A	-	D	T	C	GI 1020091	
181	S	N	D	V	C	A	Q	V	H	P	Q	K	V	T	K	F	M	L	C	A	G	R	W	T	G	G	K	S	T	C	GI 511857	
232	Q	G	D	S	G	G	P	L	V	C	N	Q	S	L	Q	G	I	I	S	W	G	Q	D	P	C	A	I	T	R	K	2723646	
209	Q	G	D	S	G	G	P	L	V	C	D	G	M	L	Q	G	I	T	S	W	G	S	D	P	C	G	K	P	E	K	GI 1020091	
211	S	G	D	S	G	G	P	L	V	C	N	G	V	L	Q	G	I	T	S	W	G	S	E	P	C	A	L	P	E	R	GI 511857	
262	P	G	V	Y	T	K	V	C	K	Y	V	D	W	I	Q	E	T	M	K	N	-	N	2723646									
239	P	G	V	Y	T	K	I	C	R	Y	T	T	W	I	K	K	T	M	D	N	R	D	GI 1020091									
241	P	S	L	Y	T	K	V	V	H	Y	R	K	W	I	K	D	T	I	V	A	N	P	GI 511857									

FIGURE 2B

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.

Y.Tom, Tang

CORLEY, Neil C.

GUEGLER, Karl J.

<120> PROSTATE-ASSOCIATED SERINE PROTEASE

<130> PF-0481 PCT

<140> To Be Assigned

<141> Herewith

<150> 09/025,059

<151> 1998-02-17

<160> 4

<170> PERL PROGRAM

<210> 1

<211> 282

<212> PRT

<213> Homo sapiens

<220> -

<223> 2723646

<400> 1

```

Met Gln Arg Leu Arg Trp Leu Arg Asp Trp Lys Ser Ser Gly Arg
 1           5           10           15
Gly Leu Thr Ala Ala Lys Glu Pro Gly Ala Arg Ser Ser Pro Leu
          20           25           30
Gln Ala Met Arg Ile Leu Gln Leu Ile Leu Leu Ala Leu Ala Thr
          35           40           45
Gly Leu Val Gly Gly Glu Thr Arg Ile Ile Lys Gly Phe Glu Cys
          50           55           60
Lys Pro His Ser Gln Pro Trp Gln Ala Ala Leu Phe Glu Lys Thr
          65           70           75
Arg Leu Leu Cys Gly Ala Thr Leu Ile Ala Pro Arg Trp Leu Leu
          80           85           90
Thr Ala Ala His Cys Leu Lys Pro Arg Tyr Ile Val His Leu Gly
          95          100          105
Gln His Asn Leu Gln Lys Glu Glu Gly Cys Glu Gln Thr Arg Thr
          110          115          120
Ala Thr Glu Ser Phe Pro His Pro Gly Phe Asn Asn Ser Leu Pro
          125          130          135
Asn Lys Asp His Arg Asn Asp Ile Met Leu Val Lys Met Ala Ser
          140          145          150
Pro Val Ser Ile Thr Trp Ala Val Arg Pro Leu Thr Leu Ser Ser
          155          160          165
Arg Cys Val Thr Ala Gly Thr Ser Cys Leu Ile Ser Gly Trp Gly
          170          175          180
Ser Thr Ser Ser Pro Gln Leu Arg Leu Pro His Thr Leu Arg Cys
          185          190          195
Ala Asn Ile Thr Ile Ile Glu His Gln Lys Cys Glu Asn Ala Tyr
          200          205          210
Pro Gly Asn Ile Thr Asp Thr Met Val Cys Ala Ser Val Gln Glu
          215          220          225
Gly Gly Lys Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val
          230          235          240
Cys Asn Gln Ser Leu Gln Gly Ile Ile Ser Trp Gly Gln Asp Pro

```

	245		250		255
Cys Ala Ile Thr	Arg Lys Pro Gly Val	Tyr Thr Lys Val Cys Lys			
	260		265		270
Tyr Val Asp Trp	Ile Gln Glu Thr Met	Lys Asn Asn			
	275		280		

<210> 2
 <211> 1314
 <212> DNA
 <213> Homo sapiens

<220> -
 <223> 2723646

<400> 2
 aacctaggcc ctcccctgcc ttgctccaca cctgggtcagg ggagagacgg gaggaagacc 60
 aaggggaagg acctaactga aaacaaacaa gctgggagaa gcaggaatct gcgctcgggt 120
 tccgcagatg cagagggtga ggtggctgcg ggactggaag tcatcgggca gaggtctcac 180
 agcagccaag gaacctgggg cccgctcctc cccctccag gccatgagga ttctgcagtt 240
 aatcctgctt gctctggcaa caggggttgt agggggagag accaggatca tcaaggggtt 300
 cgagtgaag cctcactccc agccctggca ggcagccctg ttcgagaaga cgcggctact 360
 ctgtggggcg acgctcatcg cccccagatg gctcctgaca gcagccact gcctcaagcc 420
 ccgctacata gtccacctgg ggcagcacia cctccagaag gagggagggt gtgagcagac 480
 ccggacagcc actgagtcct tccccaccc cggcttcaac aacagcctcc ccaacaaaga 540
 ccaccgcaat gacatcatgc tgggtgaagat ggcctcgcca gtctccatca cctgggctgt 600
 gcgacccctc accctctcct cagctgtgt cactgctggc accagctgcc tcatttccgg 660
 ctggggcagc acgtccagcc cccagttacg cctgcctcac accttgcat gcgccaacat 720
 caccatcatt gagcaccaga agtgtgagaa cgcctacccc ggcaacatca cagacaccat 780
 ggtgtgtgcc agcgtgcagg aagggggcaa ggactcctgc cagggtgact ccggggggccc 840
 tctgtgtgtg aaccagtctc ttcaaggcat tatctcctgg ggccaggatc cgtgtgcat 900
 caccgaaaag cctggtgtct acacgaaagt ctgcaaatat gtggactgga tccaggagac 960
 gatgaagaac aattagactg gacccaccca ccacagccca tcaccctcca tttccacttg 1020
 gtgtttggtt cctgttcaat ctgttaataa gaaaccctaa gccaaagacc tctacgaaca 1080
 ttctttgggc ctctggact acaggagatg ctgtcactta ataatacaacc tgggggttcga 1140
 aatcagtgag acctggattc aaattctgcc ttgaaatatt gtgactctgg gaatgacaac 1200
 acctgggttg ttctctgttg tatccccagc cccaaagaca gctcctggcc atatatcaag 1260
 gtttcaataa atatttgcta aatgaaaaaa aaaaaaaaaa gggcggtcgc tcta 1314

<210> 3
 <211> 260
 <212> PRT
 <213> Mus musculus

<220> -
 <223> g1020091

<400> 3
 Met Gly Arg Pro Pro Cys Ala Ile Gln Pro Trp Ile Leu Leu
 1 5 10 15
 Leu Leu Phe Met Gly Ala Trp Ala Gly Leu Thr Arg Ala Gln Gly
 20 25 30
 Ser Lys Ile Leu Glu Gly Arg Glu Cys Ile Pro His Ser Gln Pro
 35 40 45
 Trp Gln Ala Ala Leu Phe Gln Gly Glu Arg Leu Ile Cys Gly Gly
 50 55 60
 Val Leu Val Gly Asp Arg Trp Val Leu Thr Ala Ala His Cys Lys
 65 70 75
 Lys Gln Lys Tyr Ser Val Arg Leu Gly Asp His Ser Leu Gln Ser
 80 85 90
 Arg Asp Gln Pro Glu Gln Glu Ile Gln Val Ala Gln Ser Ile Gln
 95 100 105

His	Pro	Cys	Tyr	Asn	Asn	Ser	Asn	Pro	Glu	Asp	His	Ser	His	Asp
				110					115					120
Ile	Met	Leu	Ile	Arg	Leu	Gln	Asn	Ser	Ala	Asn	Leu	Gly	Asp	Lys
				125					130					135
Val	Lys	Pro	Val	Gln	Leu	Ala	Asn	Leu	Cys	Pro	Lys	Val	Gly	Gln
				140					145					150
Lys	Cys	Ile	Ile	Ser	Gly	Trp	Gly	Thr	Val	Thr	Ser	Pro	Gln	Glu
				155					160					165
Asn	Phe	Pro	Asn	Thr	Leu	Asn	Cys	Ala	Glu	Val	Lys	Ile	Tyr	Ser
				170					175					180
Gln	Asn	Lys	Cys	Glu	Arg	Ala	Tyr	Pro	Gly	Lys	Ile	Thr	Glu	Gly
				185					190					195
Met	Val	Cys	Ala	Gly	Ser	Ser	Asn	Gly	Ala	Asp	Thr	Cys	Gln	Gly
				200					205					210
Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Asp	Gly	Met	Leu	Gln	Gly	Ile
				215					220					225
Thr	Ser	Trp	Gly	Ser	Asp	Pro	Cys	Gly	Lys	Pro	Glu	Lys	Pro	Gly
				230					235					240
Val	Tyr	Thr	Lys	Ile	Cys	Arg	Tyr	Thr	Thr	Trp	Ile	Lys	Lys	Thr
				245					250					255
Met	Asp	Asn	Arg	Asp										
				260										

<210> 4

<211> 262

<212> PRT

<213> Homo sapiens

<220> -

<223> g511857

<400> 4

Met	Trp	Val	Pro	Val	Val	Phe	Leu	Thr	Leu	Ser	Val	Thr	Trp	Ile
1				5					10					15
Gly	Ala	Ala	Pro	Leu	Ile	Leu	Ser	Arg	Ile	Val	Gly	Gly	Trp	Glu
				20					25					30
Cys	Glu	Lys	His	Ser	Gln	Pro	Trp	Gln	Val	Leu	Val	Ala	Ser	Arg
				35					40					45
Gly	Arg	Ala	Val	Cys	Gly	Gly	Val	Leu	Val	His	Pro	Gln	Trp	Val
				50					55					60
Leu	Thr	Ala	Ala	His	Cys	Ile	Arg	Lys	Cys	Lys	Ser	Val	Ile	Leu
				65					70					75
Leu	Gly	Arg	His	Ser	Leu	Phe	His	Pro	Glu	Asp	Thr	Gly	Gln	Val
				80					85					90
Phe	Gln	Val	Ser	His	Ser	Phe	Pro	His	Pro	Leu	Tyr	Asp	Met	Ser
				95					100					105
Leu	Leu	Lys	Asn	Arg	Phe	Leu	Arg	Pro	Gly	Asp	Asp	Ser	Ser	His
				110					115					120
Asp	Leu	Met	Leu	Leu	Arg	Leu	Ser	Glu	Pro	Ala	Glu	Leu	Thr	Asp
				125					130					135
Ala	Val	Lys	Val	Met	Asp	Leu	Pro	Thr	Gln	Glu	Pro	Ala	Leu	Gly
				140					145					150
Thr	Thr	Cys	Tyr	Ala	Ser	Gly	Trp	Gly	Ser	Ile	Glu	Pro	Glu	Glu
				155					160					165
Phe	Leu	Thr	Pro	Lys	Lys	Leu	Gln	Cys	Val	Asp	Leu	His	Val	Ile
				170					175					180
Ser	Asn	Asp	Val	Cys	Ala	Gln	Val	His	Pro	Gln	Lys	Val	Thr	Lys
				185					190					195
Phe	Met	Leu	Cys	Ala	Gly	Arg	Trp	Thr	Gly	Gly	Lys	Ser	Thr	Cys
				200					205					210
Ser	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Asn	Gly	Val	Leu	Gln
				215					220					225
Gly	Ile	Thr	Ser	Trp	Gly	Ser	Glu	Pro	Cys	Ala	Leu	Pro	Glu	Arg

	230		235		240
Pro Ser Leu Tyr	Thr Lys Val Val His	Tyr Arg Lys Trp Ile	Lys		
	245	250	255		
Asp Thr Ile Val	Ala Asn Pro				
	260				